

**TransMiner: Mining Transitive Associations Among Biological Objects From Text**  
**Narayanasamy, Vijay<sup>\*1</sup>, Mukhopadhyay, Snehasis<sup>2</sup>, Palakal, Mathew<sup>2</sup>, Potter, David<sup>3</sup>**

**<sup>1</sup>School of Informatics; <sup>2</sup>Department of Computer and Information Science; <sup>3</sup>Indiana University School of Medicine, Indiana University-Purdue University Indianapolis, Indianapolis, IN, USA**

**Summary:**

Associations among biological objects such as genes, proteins, and drugs can be discovered automatically from scientific literature. 'TransMiner' is a system for finding associations among objects by mining the PubMed database of scientific literature. The direct associations among the objects are discovered based on the principle of co-occurrence from the text documents in the form of an association graph. The principle of transitive closure is applied to the association graph to find potential transitive associations. The potential transitive associations that are indeed direct are discovered by iterative retrieval and mining of the PubMed documents. Those associations that are not found explicitly in the entire PubMed database are transitive associations and are the candidates for hypotheses generation. All the discovered direct associations were manually evaluated. The transitive associations were ranked based on the sum of weight of terms that co-occur with both the biological objects. The direct and transitive associations are visualized using a graph visualization applet for use by the scientists.

**Results:**

TransMiner was used to identify associations among 24 objects in the areas of diabetes and cell motility/cytoskeletal remodeling with the help of user defined dictionary. The dictionary terms include known calpain substrates and the signaling proteins that are associated either upstream or downstream of calpain. Out of 178 direct associations discovered by TransMiner 170 direct associations (95.55%) were found to have some valid biological association. From the discovered direct associations, EGFR-mediated calpain activation, calpain activation of NF- $\kappa$ B signaling and SNAP/SNARE mediated regulation of secretion were identified as active areas of investigation by filtering the associations using the graph visualization applet. The transitive associations were found most strongly in the area of SNAP/SNARE mediated secretion. TransMiner was therefore able to identify the most actively investigated areas and to point to the areas with potential for future discovery.

**Availability:**

Graph visualization applets and result tables are available at <http://sifter.cs.iupui.edu/~sifter/transMiner>

*National Science Foundation Information Technology Research: NSF-IIS/ITR 0081944*

*National Institutes of Health: NIH-NIGMS P20-GM66402-01*